

Aplikasi Teori Respon Butir Untuk Menguji Invariansi Pengukuran Psikologi Guna Keperluan Survei dan Seleksi Pekerjaan

Wahyu Widhiarso
Fakultas Psikologi, Universitas Gadjah Mada

This study aims was exploring model that explain faking responses on the psychological scale through item response theory modeling (IRT) and tested its goodness of fit. This research use quasi-experimental design that manipulates participant instructions to response Five Factors Personality Scale. In the first instruction, participants were asked to give honest response to given scale, while the second condition were asked to imagine as job applicants so they consider to manipulate their response to give positive impression. IRT based modeling was done on both types of response by using partial credit model (PCM). Results of analysis suggest whether there is no difference in threshold parameters between the two types of responses. Additional analysis with differential item functioning (DIF) test found inconsistent results. There are some items affected by DIF that indicated participants who manipulated their response have probability of getting a high score than others. Overall, this study shows that the nature invariance of IRT modeling interrupted by faking response.

Keywords: item response theory, faking response, differential item functioning

Penelitian banyak yang telah membuktikan bahwa skala psikologi sangat rentan terhadap respons tipuan sehingga diperlukan penelitian yang intensif untuk mengatasi permasalahan ini (Zickar & Robie, 1999). Viswesvaran dan Ones (1999) melalui studi meta analisis menemukan bahwa rerata skor skala kepribadian dari aplikasi yang mengikuti seleksi kerja lebih tinggi 0,48 hingga 0,65 di atas rerata skor dari subjek yang telah bekerja. Penelitian lain juga menunjukkan bahwa hadirnya respons tipuan mempengaruhi skor total (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), estimasi

theta dengan menggunakan IRT (Zickar & Drasgow, 1996), struktur faktor tes kepribadian baterai (Schmit & Ryan, 1993), korelasi antar sub-skala (Douglas, McDaniel, & Snell, 1996) dan validitas kriteria dan utilitas sistem seleksi (Zickar, Rosse, & Levin, 1996). Paparan tersebut menunjukkan bahwa pengukuran psikologi rentan terhadap respon yang menipu dan adanya respon menipu tersebut dapat mengganggu properti psikometris pengukuran yang dilakukan.

Penelitian mengenai respons tipuan (*faking*) telah banyak dilakukan dengan desain

penelitian yang dipakai sangat beragam. Pertama, beberapa peneliti berusaha mengidentifikasi tipuan dengan bantuan instrumen pengukuran yang khusus mendeteksi respons tipuan misalnya *Balanced Inventory of Desirable Responding* (Paulhus, 1984), *Unlikely Virtues Scales, Assessment of Background and Life Experiences* (ABLE), *Validity Scale*, serta Skala Kecenderungan Sosial (*Social Desirability Scale*) yang dikembangkan oleh beberapa peneliti, misalnya *Social Desirability Scale* (Edwards, 1957), *Marlowe-Crowne Social Desirability Scale* (Crowne & Marlowe, 1960), dan *Jackson Social Desirability Scale*. Zickar dan Robie (1999) mengatakan bahwa meskipun skala tersebut didesain untuk tahan terhadap respons tipuan, hasil penelitian menunjukkan bahwa banyak individu yang bisa mengatasi 'jebakan' ini. Korelasi antara skala tersebut dengan skala-skala kepribadian yang bersumber dari aplikasi pekerjaan lebih tinggi dibanding dari responden non-aplikan. Dapat disimpulkan bahwa instrumen-instrumen tersebut tidak tahan terhadap respons tipuan.

Kedua, desain penelitian dalam mengidentifikasi respons tipuan dilakukan dengan membandingkan rerata skor total skala psikologi dari dua kelompok yang memiliki motivasi yang berbeda, misalnya subjek pelamar pekerjaan dibandingkan dengan subjek penelitian atau subjek yang diinstruksikan menjawab jujur dan subjek yang diinstruksikan memberikan respon yang menggambarkan citra diri sebaik-baiknya. Desain ini memiliki keterbatasan karena fenomena respons tipuan tidak berada pada tataran skor total akan tetapi pada tataran skor butir. Von Davier (2010) mengatakan bahwa skor total kurang memberikan banyak informasi karena tidak bisa

menjangkau dinamika psikologis secara mendetail. Dalam konteks penelitian mengenai respons tipuan, investigasi melalui skor total juga kurang direkomendasikan (Zickar & Gibby, 2006).

Ketiga, respons tipuan dikaji melalui analisis faktor baik eksploratori maupun konfirmatori (Montag & Comrey, 1990). Perbedaan struktur faktor data yang didapatkan dari partisipan yang memiliki motivasi berbeda (netral vs. motivatif) diinterpretasikan merupakan resultan dari respons tipuan. Asumsi bahwa kontrak empirik berkaitan dengan faktor secara linier menyebabkan teknik analisis ini kurang berhasil jika diterapkan pada skala psikologi yang menyediakan alternatif respon yang sedikit, misalnya dua alternative respons (ya dan tidak). Ditambah lagi dengan ukuran sampel yang sedikit dan jumlah faktor yang diekstrak terlalu banyak akan menghambat kestabilan hasil analisis faktor yang diterapkan (S. Stark, O. S. Chernyshenko, K. Y. Chan, W. C. Lee, & F. Drasgow, 2001).

Keempat, respons tipuan dikaji melalui teori respon butir (*Item Response Theory/IRT*) (Zickar & Robie, 1999). Teknik IRT yang banyak digunakan adalah perbedaan keberfungsian butir (*differential item functioning/DIF*) yang dipakai mengidentifikasi perbedaan seberapa jauh probabilitas kelompok referen (respon jujur) dan fokal (respon menipu) dalam mengatasi butir dalam skala psikologi (Stark et al., 2001). Perbedaan keberfungsian tes (*differential test functioning/DTF*) juga dipakai untuk melihat apakah properti pengukuran satu kelompok dengan kelompok lainnya memiliki perbedaan setelah rerata perbedaan skor murni dikendalikan (Woods, Oltmanns, & Turkheimer, 2008).

Secara metodologis IRT merupakan pelengkap dari teknik analisis konfirmatori (Reise, Widaman, & Pugh, 1993). Berbeda dengan analisis faktor konfirmatori yang dipakai untuk mengidentifikasi hubungan yang linier antara respon terhadap indikator, IRT dapat dipakai untuk mengidentifikasi hubungan antara pola respons individu terhadap indikator dengan faktor ukurnya. Selain itu kelebihan IRT dibanding analisis faktor konfirmatori adalah tersedianya model-model alternatif pilihan respon, misalnya pada model politomi dikenal model parsial kredit (PCM), skala rating (RM) dan model respons bergradasi (GRM). Model politomi memiliki keunggulan karena dapat dipakai untuk menyusun model yang menjekaskan interaksi antara subjek dengan butir. Oleh karena itu model politomi dapat diterapkan pada skala kepribadian dengan format Likert yang memiliki alternatif respon bergradasi dari setuju hingga tidak setuju. Model IRT politomi yang banyak dipakai untuk mengidentifikasi skor dari Skala Likert adalah model response bergradasi (Samejima, 1969) dan model kredit parsial (Masters, 1982), namun penelitian ini menggunakan model kredit parsial karena pendekatan yang dipakai dalam penelitian ini pendekatan Rasch.

Pemodelan Respons Tipuan

Menurut Zickar & Robie (1999) pemodelan respons tipuan dapat dibagi menjadi dua jenis, yaitu model perubahan butir dan perubahan individu. Model perubahan butir mengasumsikan bahwa respons tipuan menyebabkan perubahan parameter butir dan alternatif pilihan respon. Karena memiliki harapan atau intensi yang berbeda, maka individu yang merespon jujur dan menipu pada

butir yang memiliki struktur yang sama akan memiliki persepsi yang berbeda terhadap butir yang dihadapi. Dengan demikian, karena perubahan hanya terletak pada parameter model maka nilai theta yang menunjukkan abilitas kedua orang di atas tidak berbeda.

Model perubahan individu menunjukkan bahwa respons tipuan menyebabkan informasi mengenai abilitas laten individu menjadi berbeda. Subjek yang merespon butir dengan tidak jujur akan memperoleh nilai theta yang tinggi. Model ini bertolak belakang dengan asumsi IRT yang menjelaskan bahwa theta relatif stabil dan tidak terpengaruh oleh karakteristik subjek dan situasi pengukuran. Individu yang melakukan penipuan respon diasumsikan akan merespon butir seakan-akan dia memiliki nilai theta yang tinggi atau rendah dari nilai theta dia sebenarnya. Asumsi tersebut dapat diartikan bahwa model ini memperkenankan parameter butir bervariasi. Dengan mengasumsikan parameter butir bervariasi maka perubahan nilai theta tidak bersifat konstan, akan tetapi berbeda-beda pada tiap butir (Zickar dan Robie, 1999).

Beberapa ahli telah mengembangkan model dalam menggambarkan respon menipu pada level butir. Levine dan Rubin (1979) mengembangkan model yang dinamakan dengan ketepatan pengukuran (*appropriateness measurement*) yang dapat dipakai untuk menguji apakah respon individu telah sesuai dengan model IRT yang menggambarkan pola respon individu di dalam sampel. Model tersebut dapat mengidentifikasi perbedaan antar pola respon individu teramati dengan pola respon individu harapan berdasarkan posisi individu pada rentang nilai theta dan perangkat fungsi respon butir (IRF). Fungsi respon butir adalah fungsi yang menunjukkan hubungan

posisi trait individu dengan probabilitas dalam merespon butir.

Beberapa peneliti telah mengembangkan model untuk mendeteksi kemunculan dan kuantitas respons tipuan. Levine dan Drasgow (1988) mengembangkan ketepatan pengukuran optimal yang memadukan antara respon netral dengan respon tipuan yang dapat dipakai untuk mendeteksi respon menipu. Zickar dan Drasgow (1996) mengembangkan model respon menipu yang dinamakan dengan model perpindahan nilai theta (*theta-shift models*) yang mendeteksi perubahan nilai theta akibat respons tipuan. Model tersebut merupakan modifikasi dari model dua parameter (IRT 2-PL) yang banyak dipakai dalam mengidentifikasi properti data dari skala kepribadian dan skala sikap karena model ini cukup sederhana dan memiliki properti yang cukup atraktif. Peneliti banyak membuktikan ketepatan model ini dengan data yang berasal pengukuran kepribadian (Reise & Waller, 1990). Model perpindahan nilai theta mengasumsikan pada butir yang diduga rentan terhadap respon menipu, individu yang memberikan respon menipu akan mengalami peningkatan nilai theta (misalnya sebesar +0,5) pada kontinum skala theta.

Model IRT politomi dasar juga dapat dipakai dalam pemodelan respon menipu. Sub-model IRT politomi yang banyak dipakai adalah model GRM. Zickar dan Robie (1999) melakukan pemodelan respons tipuan dengan model GRM dari Samejima dengan menekankan pada ketepatan data dengan fungsi pilihan respons (*option respon function/ORF*). Ketepatan model ditentukan melalui uji kai-kuadrat yang menguji perbedaan antara jumlah respon pilihan harapan dan jumlah respon pilihan dari data. Kai-kuadrat dihitung

berdasarkan jumlah waktu harapan bahwa responden akan memilih pilihan- k pada model probabilitas IRT. Tingginya nilai kai-kuadrat menunjukkan bahwa model tidak tepat dengan data.

Penelitian mengenai pemodelan respon menipu pada skala kepribadian melalui pemodelan IRT memberikan manfaat yang besar dalam pengembangan skala psikologi. Informasi mengenai pengaruh respons tipuan terhadap parameter butir akan memudahkan peneliti untuk memahami invariansi properti psikometris butir. Informasi ini sekaligus akan menjadi masukan bagi pengembang alat ukur psikologi dalam menyusun butir yang tahan terhadap respons tipuan.

Model yang akurat mengenai respon yang menipu dapat dijadikan sebagai panduan untuk mengidentifikasi respon tipuan subjek terhadap skala kepribadian. Model IRT yang dipakai dalam penelitian ini adalah model kredit parsial (PCM).

Penelitian ini memiliki dua tujuan, pertama untuk mengeksplorasi pengaruh respon menipu terhadap perubahan parameter butir melalui pemodelan IRT. Investigasi pengaruh respons tipuan terhadap parameter butir dapat dilakukan dengan membandingkan parameter butir dari dua kondisi pengukuran, yaitu kondisi jujur dan kondisi menipu. Kedua, mengidentifikasi model IRT yang tepat dalam menggambarkan respons tipuan.

Metode

Desain

Penelitian ini menggunakan metode eksperimen kuasi dengan dua kondisi perlakuan. Pada kondisi pertama (kondisi

netral), peneliti menginstruksikan subjek untuk merespon butir dengan jujur, pada kondisi kedua (kondisi motivatif) peneliti menginstruksikan subjek untuk merespon butir seakan-akan sebagai aplikasi yang sedang melamar pekerjaan dan sedang mengikuti proses seleksi pekerjaan. Untuk menghindari efek urutan, peneliti membagi subjek menjadi dua kelompok dengan urutan pemberian instruksi yang berbeda.

Partisipan

Partisipan yang berpartisipasi dalam eksperimen adalah mahasiswa yang Fakultas Psikologi UGM yang berjumlah 120 (46% laki-laki dan 54% perempuan). Usia sampel bergerak dari 19-23 tahun ($M = 20,21$). Pelaksanaan eksperimen dilakukan di kelas setelah mereka mengikuti perkuliahan. Sebelum melakukan eksperimen peneliti memperkenalkan diri dan menjelaskan tujuan kegiatan yang dilakukan. Semua mahasiswa setuju untuk mengisi lembar persetujuan partisipasi sebelum mereka mengikuti jalannya eksperimen dengan mengisi instrumen yang dibagikan.

Instrumen

Instrumen yang dipakai untuk mengukur adalah Inventori Kepribadian Lima Faktor yang diadaptasi oleh peneliti dari *Big Five Inventory (BFI)* yang dikembangkan oleh John, Donahue, & Kentle (1991). Instrumen ini menggunakan model skala Likert yang terdiri dari lima alternatif respons yang berbenyuk pelaporan mandiri (*self report*). Subjek diminta untuk melengkapi butir-butir pernyataan yang menggambarkan berbagai karakteristik individu. Respon yang disediakan ada lima

alternatif respons dari sangat setuju hingga sangat tidak setuju dengan penyekoran butir bergerak dari 1 hingga 5.

Skala ini mengukur lima faktor kepribadian antara lain ekstreversi (*extroversion*), keramahan (*agreeableness*), ketelitian (*conscientiousness*), neurotisme (*neuroticism*) dan keterbukaan (*openness*). Dari kelima faktor BFI, hanya tiga faktor yang dilibatkan dalam penelitian ini, yaitu faktor kepribadian teliti, ekstrovert dan terbuka. Ketiga faktor tersebut dalam penelitian-penelitian sebelumnya lebih terkait dengan pekerjaan dibanding dengan dua faktor lainnya (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). BFI versi Bahasa Indonesia telah diujicobakan oleh peneliti pada sampel mahasiswa ($n = 185$) yang menghasilkan nilai reliabilitas (α) sebagai berikut: faktor ekstreversi (0,839), keramahan (0,789), ketelitian (0,924), kestabilan emosi (0,848) dan keterbukaan (0,807). Hasil ini mirip versi asli yang dilaporkan oleh John dan Srivastava BFI memiliki reliabilitas (α) antara 0,75 hingga 0,80 dan reliabilities tes-tes ulang antara 0,80 hingga 0,90. Validitas konkuren BFI cukup tinggi yang terlihat dari korelasi yang tinggi dengan NEO-FFI dan TDA menghasilkan rata-rata korelasi sebesar 0,83 hingga 0,91 (John & Srivastava, 1999).

Analisis

Analisis data dilakukan dengan menggunakan prosedur pemodelan IRT. Prosedur yang dilakukan adalah pengujian invariansi dan unidimensionalitas model yang kemudian dilanjutkan pada kalibrasi parameter butir. Proses kalibrasi ini dilakukan dengan menggunakan model kredit parsial (PCM) dari Master Masters (1982). Model ini termasuk

dalam pendekatan model pengukuran Rasch yang menitikberatkan pada lokasi butir pada proses pemodelannya. Prosedur ini dilakukan pada data dari kondisi netral saja karena model ini dipakai untuk mengevaluasi model yang didapatkan dari data motivatif. Untuk menguji perbedaan model tersebut peneliti melakukan uji perbedaan keberfungsian butir (DIF) antar dua kondisi. Program lunak komputer yang dipakai untuk melakukan prosedur pemodelan ini adalah Winstep (Linacre, 2000).

Hasil

Deskripsi data

Deskripsi data statistik dapat dilihat pada Tabel 1 yang menunjukkan rerata (M) dan deviasi standar (SD) pada masing-masing faktor kepribadian dan kondisi pengukuran. Secara umum rerata antara dua kondisi pengukuran berbeda yang terlihat dari tingginya rerata skor faktor kepribadian pada kondisi motivatif dibanding dengan kondisi netral. Hal ini menunjukkan bahwa partisipan mampu meningkatkan nilai skor skalanya ketika diminta untuk dengan sengaja memanipulasi tanggapannya terhadap butir-butir skala. Nilai deviasi standar antar dua kondisi tidak berbeda jauh yang menunjukkan adanya peningkatan sistematis skor skala subjek dari kondisi netral menuju kondisi motivatif. Dengan kata lain, sebagian besar peningkatan skor partisipan meningkat dalam selisih skor yang sama antar dua kondisi. Deskripsi statistik pada tataran butir dapat dilihat pada Tabel 2 yang menunjukkan rerata dan deviasi standar skor butir pada tiap faktor kepribadian.

Variasi asumsi model

Teori respon butir memiliki beberapa asumsi yang perlu diverifikasi sebelum proses pemodelan dilakukan. Asumsi tersebut adalah unidimensionalitas data, independensi lokal dan invariansi pengukuran (Embretson & Reise, 2000). Unidimensionalitas menunjukkan apakah model mengukur atribut tunggal ataukah majemuk, independensi lokal menunjukkan apakah respons terhadap satu butir tidak dipengaruhi oleh respons terhadap butir lainnya sedangkan invariansi menunjukkan bahwa model tersebut tidak dipengaruhi oleh karakteristik sampel.

Konfirmasi asumsi pada penelitian ini dilakukan dengan prosedur berikut: unidimensionalitas data diuji dengan menggunakan analisis faktor konfirmatori, independensi lokal ditunjukkan dengan penelaahan butir-butir yang tidak tumpang tindih dalam mengukur target ukur, dan invariansi dilakukan dengan mengidentifikasi kesamaan hasil kalibrasi parameter butir. Invariansi dibuktikan dengan tingginya korelasi antara hasil kalibrasi parameter butir antara kelompok sampel yang dibagi secara acak. Prosedur konfirmasi ini dilakukan pada data dari kondisi netral.

Hasil analisis data menunjukkan bahwa semua asumsi IRT pada semua faktor kepribadian dapat dibuktikan. Hasil analisis dimensionalitas pengukuran melalui analisis faktor konfirmatori menunjukkan bahwa asumsi model pengukuran tunggal pada masing-masing konstruk menghasilkan indeks ketepatan model yang dapat dibuktikan. Hasil analisis invariansi menunjukkan bahwa semua data bersifat invarian. Pengujian invariansi menunjukkan bahwa korelasi antara hasil kalibrasi pada kedua

kelompok semuanya tinggi (0,76 – 0,89). Pengujian DIF yang dilakukan sebagai uji pelengkap pengujian invariansi menunjukkan bahwa terdapat satu butir yang memiliki perbedaan hasil kalibrasi parameter, yaitu satu pada faktor keramahan dan satu pada faktor emosi stabil. Hasil ini tidak mempengaruhi kesimpulan peneliti karena penelitian lebih mengutamakan pada korelasi antar hasil kalibrasi parameter.

Pemodelan IRT

Hasil pemodelan IRT pada tiap faktor disajikan pada Tabel 4 yang menunjukkan perbandingan parameter butir ditinjau berdasarkan kondisi pengukuran dan faktor kepribadian. Statistik yang dipaparkan pada tabel tersebut adalah parameter ambang batas (*threshold*) yang menunjukkan titik pada sepanjang kurva karakteristik butir (ICC), yang memperlihatkan probabilitas respons untuk mendapatkan skor yang tinggi pada butir tertentu. Semakin besar nilai parameter ini semakin mudah seorang individu untuk mendapatkan skor tinggi. Dalam tes kognitif, parameter ini dinamai tingkat kesukaran butir yang menunjukkan semakin tinggi tingkat kesukaran butir maka probabilitas subjek untuk mendapatkan skor tinggi semakin rendah.

Parameter lainnya yang dipaparkan adalah OUTFIT dan INFIT beserta nilai standarnya (ZSTD). OUTFIT adalah statistik yang menunjukkan informasi yang menunjukkan kecocokan antara butir dan subjek yang tidak terstandarisasi dan terbobot. INFIT adalah statistik yang menunjukkan informasi yang menunjukkan kecocokan antara butir dan subjek yang terstandarisasi dan terbobot. Ketika data sesuai dengan model, maka statistik ini nilainya

akan mendekati nilai t yang dalam pengaturan ini, perkiraan nilai t memiliki rerata sebesar 0 dan deviasi standar sebesar 1.

Hasil analisis yang ditunjukkan oleh Tabel 4 menginformasikan bahwa hasil pengujian parameter butir baik pada kondisi netral dan kondisi motivasional tidak memiliki perbedaan. Nilai ambang untuk tiap kondisi adalah sama. Sebagian besar parameter butir pada INFIT dan OUTFIT adalah setara. Nilai rerata Z pada INFIT dan OUTFIT untuk antar dua kondisi adalah setara. Diantara kelima faktor kepribadian, faktor keterbukaan memiliki nilai kecocokan butir yang rendah yang terlihat pada tingginya nilai deviasi standar INFIT dan OUTFIT baik pada kondisi netral dan kondisi motivatif. Hal ini menunjukkan bahwa rata-rata butir-butir dalam mengukur keterbukaan kurang tepat dikenakan pada sampel penelitian. Permasalahan ini tidak dibahas lebih lanjut dalam penelitian ini karena fokus penelitian ini adalah membandingkan hasil pemodelan antara dua kondisi pengukuran. Kesimpulan analisis yang didapatkan adalah pemodelan pada respons netral dan motivatif secara umum memiliki kesetaraan.

Pengujian Diferensial Keberfungsian Butir

Selain membandingkan pemodelan berdasarkan IRT, penelitian ini juga bertujuan untuk membandingkan probabilitas individu untuk mendapatkan skor tinggi pada pengukuran kepribadian dengan menggunakan teknik diferensial keberfungsian butir (DIF). Hasil pengujian dapat dilihat pada Tabel 5. Tabel tersebut menunjukkan bahwa pada faktor ekstraversi dari 9 butir yang mengukurnya, 8 butir terjangkau DIF. Hasil tersebut menunjukkan bahwa probabilitas partisipan

Tabel 1.
Deskripsi Statistik Skor Skala Pada Tiap Faktor Kepribadian dan Kondisi Pengukuran

Faktor	Kondisi	M	SD	SE	Skewness	Kurtosis
Ekstraversi	Netral	31,832	5,951	0,611	-0,174	-0,114
	Motivatif	37,044	4,432	0,467	-0,321	0,520
Keramahan	Netral	35,337	3,962	0,406	-0,323	0,453
	Motivatif	39,483	3,721	0,394	-0,096	-1,047
Keuletan	Netral	30,495	5,945	0,610	0,030	-0,273
	Motivatif	38,767	5,208	0,549	-1,331	2,763
Emosi Stabil	Netral	24,400	4,483	0,460	0,089	1,033
	Motivatif	30,444	4,037	0,426	0,102	-0,104
Keterbukaan	Netral	33,200	5,400	0,554	-0,014	0,619
	Motivatif	37,878	4,434	0,467	-0,319	0,248

Tabel 2.
Deskripsi Statistik Skor Butir Pada Tiap Faktor Kepribadian

Nomor butir	Faktor kepribadian				
	Ekstraversi	Keramahan	Keuletan	Emosi Stabil	Keterbukaan
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
1	3,87 (1,01)	4,00 (0,85)	4,23 (0,72)	3,76 (0,88)	3,94 (0,89)
2	3,73 (1,10)	3,91 (0,87)	3,92 (1,14)	3,91 (0,69)	4,37 (0,63)
3	3,79 (0,93)	4,36 (0,64)	4,11 (0,83)	2,62 (0,95)	3,98 (0,81)
4	4,10 (0,79)	4,30 (0,75)	3,62 (1,18)	3,54 (1,02)	4,04 (0,81)
5	3,58 (1,14)	4,35 (0,65)	3,68 (1,17)	3,67 (0,95)	3,99 (0,86)
6	3,68 (0,98)	4,21 (0,83)	3,64 (0,97)	2,80 (1,07)	4,22 (0,67)
7	3,76 (1,08)	4,22 (0,62)	3,94 (0,85)	3,59 (0,95)	3,35 (1,06)
8	4,01 (0,85)	3,73 (1,10)	4,01 (0,78)	3,45 (0,95)	3,89 (0,82)
9	3,85 (0,94)	4,27 (0,75)	3,37 (1,04)		3,69 (2,27)

Tabel 3.
Hasil Pengujian Invariansi Estimasi Parameter Model Pada Tiap Faktor Kepribadian

Faktor Kepribadian	Unidimensi	Invariansi	RMSEA	r_{xy}	ΣDIF
	λ^2 (db = 24)	GFI			
Ekstraversi	42,16	0,92	0,08	0,86	0
Keramahan	25,51	0,94	0,02	0,72	1
Keuletan	36,18	0,92	0,08	0,89	0
Emosi stabil	32,73	0,93	0,07	0,76	1
Keterbukaan	30,25	0,93	0,04	0,85	0

Keterangan. Nilai kai kuadrat unidimensi faktor ekstraversi memiliki signifikansi di atas 0.05 namun memiliki nilai indeks pendukung yang cukup tinggi untuk penerimaan model yaitu GFI (0.92) dan RMSEA (0.08)

Tabel 4.
Deskripsi Statistik Skor Butir pada Tiap Faktor Kepribadian

Kondisi	Faktor Kepribadian	Nilai Ambang		Infit		Outfit	
		<i>M</i>	<i>SE</i>	<i>MNSQ</i>	<i>Z.STD</i>	<i>MNSQ</i>	<i>Z</i>
Netral	Ekstraversi	0,00 (0,59)	0,15 (0,02)	1,01 (0,27)	0,00 (1,70)	1,03 (0,31)	0,00 (1,90)
	Keramahan	0,00 (0,53)	0,16 (0,02)	1,01 (0,12)	0,10 (0,80)	1,00 (0,14)	0,00 (0,90)
	Keuletan	0,00 (0,76)	0,17 (0,02)	0,98 (0,19)	-0,10 (1,40)	1,06 (0,30)	0,30 (1,90)
	Emosi Stabil	0,00 (0,42)	0,16 (0,02)	0,99 (0,19)	-0,10 (1,50)	0,99 (0,18)	-0,10 (1,40)
	Keterbukaan	0,00 (1,03)	0,15 (0,04)	0,92 (0,51)	-0,50 (2,90)	1,05 (0,50)	0,00 (3,00)
Motivatif	Ekstraversi	0,00 (0,48)	0,17 (0,04)	0,99 (0,17)	-0,20 (1,00)	1,12 (0,30)	0,40 (1,50)
	Keramahan	0,00 (0,69)	0,20 (0,04)	0,99 (0,10)	-0,10 (0,50)	1,09 (0,35)	0,30 (1,30)
	Keuletan	0,00 (0,64)	0,22 (0,02)	1,01 (0,45)	-0,30 (1,90)	1,00 (0,52)	-0,20 (1,90)
	Emosi Stabil	0,00 (0,39)	0,16 (0,02)	0,99 (0,26)	-0,20 (1,70)	1,05 (0,38)	0,10 (2,30)
	Keterbukaan	0,00 (0,67)	0,21 (0,03)	0,96 (0,49)	-0,50 (2,50)	1,03 (0,70)	-0,20 (3,10)

Keterangan. Yang ditampilkan pada tabel adalah rerata parameter dari butir-butir di dalam faktor. Angka yang di dalam kurung adalah deviasi standar dari parameter tersebut.

Tabel 5.
Hasil Uji Perbedaan Keberfungsian Butir Diferensial Pada Tiap Faktor Kepribadian

Nomor Butir	Faktor kepribadian				
	Ekstraversi	Keramahan	Keuletan	Emosi stabil	Keterbukaan
a1	3,298**	10,242**	0,196	0,150	4,948*
a2	26,280**	4,799*	3,254	2,494	0,465
a3	9,567**	0,900	0,467	0,491	2,124
a4	2,265	1,927	1,048	0,267	0,000
a5	10,164**	1,158	0,098	0,092	5,755*
a6	10,010**	6,646**	0,272	0,485	11,915**
a7	38,879**	2,841	0,784	2,055	0,282
a8	43,701**	4,830*	3,247	0,246	3,518
a9	21,566**	0,255	1,151		7,799**
a10					

Keterangan. Yang ditampilkan adalah hasil nilai kai-kuadrat.

** $p < 0,01$ * $p < 0,05$

untuk mendapatkan skor tinggi pada kondisi motivatif lebih tinggi dibanding ada kondisi netral. Dengan kata lain, partisipan pada kondisi motivatif cenderung untuk memberikan respon yang mendukung tingginya perolehan skor pada faktor ini. Selain faktor ekstraversi, faktor-faktor lain menghasilkan adanya butir yang terjangkit DIF yaitu faktor keramahan (4 butir) dan keterbukaan (3 butir). Butir-butir pada faktor keuletan dan emosi stabil tidak ada yang terjangkit DIF.

Diskusi

Penelitian ini menghasilkan beberapa temuan. *Pertama*, pemodelan IRT dengan menggunakan data yang berasal dari data partisipan yang memanipulasi responnya terhadap butir skala secara umum memiliki kesamaan dengan data partisipan yang tidak memanipulasi responnya. Hasil kalibrasi parameter butir pada kedua kondisi menunjukkan nilai ambang yang setara pada semua faktor skala kepribadian lima faktor.

Kedua, probabilitas untuk mendapatkan skor butir yang tinggi antara partisipan yang netral dan partisipan memiliki motivasi (memberikan impresi positif), cenderung bervariasi. Pada faktor keterbukaan, keramahan dan keterbukaan, beberapa butir memberikan probabilitas yang berbeda antara kedua jenis respons tersebut. Partisipan yang memanipulasi responsnya memiliki probabilitas yang tinggi untuk mendapatkan skor yang tinggi pada butir skala. Sebaliknya pada faktor keuletan dan emosi stabil probabilitas antar kedua jenis respons tersebut adalah sama.

Temuan pertama penelitian ini membuktikan asumsi pemodelan IRT yang

menyatakan bahwa pemodelan IRT adalah sampel independen. Artinya nilai parameter butir di dalam model tidak terpengaruh oleh karakteristik sampel (invarian). Selama ini penelitian mengenai independensi pemodelan sampel IRT dipakai terhadap karakteristik sampel dilakukan berdasarkan karakteristik sampel berdasarkan kondisi demografis (misalnya jenis kelamin dan budaya) maupun karakteristik sampel (misalnya tingkat kecerdasan dan unit kelompok). Penelitian ini menambahkan informasi baru bahwa pemodelan IRT ternyata juga invarian terhadap sampel yang netral dan sampel yang memiliki motivasi tertentu.

Tidak adanya perbedaan antara model IRT antara partisipan yang jujur dan menipu dapat disebabkan oleh pola respons partisipan ketika pada situasi netral dan situasi motivatif adalah sama. Peningkatan terjadi hanya pada nilai rerata skor saja akan tetapi tidak pada nilai varians. Peningkatan rerata ini menyebabkan adanya perbedaan yang terjadi adalah perbedaan sistematis. Dengan adanya perbedaan yang sistematis antara skor dari kondisi jujur dan menipu menyebabkan analisis tidak mendeteksi adanya perbedaan model. Kesamaan parameter model tersebut dikarenakan nilai varians respons pada kedua kondisi tersebut adalah sama. Variabilitas yang sama ini mengurangi kemanjuran pendeteksian terhadap respons yang khas dan unik yang diberikan oleh partisipan yang menipu. Ketika respon butir yang unik tersebut digabungkan dengan respon butir yang unik secara keseluruhan, maka keunikan tersebut saling membatalkan satu dengan lainnya. Akibatnya model IRT yang dihasilkan konstan dengan model IRT pada sampel jujur.

Invariansi berdasarkan karakteristik sampel

dari fitur parameter butir merupakan salah satu kekuatan pemodelan IRT. Pernyataan bahwa nilai-nilai parameter butir bukan hanya untuk kelompok tertentu yang menanggapi butir (Natarajan, 2009) telah dibuktikan dalam penelitian ini. Parameter butir yang diperkirakan dari setiap segmen kurva respons butir menunjukkan bahwa parameter tersebut dapat diperkirakan dari setiap kelompok pengambil tes. Dengan adanya berlakunya invariansi pemodelan IRT berdasarkan situasi yang dihadapi partisipan, maka pemodelan khusus untuk partisipan yang memberikan respon menipu tidak dapat dikenakan. Hal ini dikarenakan hasil pemodelan sampel yang melakukan tipuan akan sama dengan sampel yang memberikan respon jujur. Dengan tidak berlakunya pemodelan khusus untuk partisipan yang melakukan tipuan maka estimasi nilai theta tidak dapat dipakai untuk membedakan mana partisipan yang jujur dan menipu. Hal ini sama halnya dengan membedakan skor skala yang tinggi dan rendah yang tidak dapat dipakai untuk mengetahui tingginya skor skala tersebut dikarenakan mereka menipu ataukah tidak. Fenomena ini telah diperkirakan oleh Zickar dan Robie (1999) yang menggunakan istilah perpindahan nilai theta (*theta shift*).

Penyebab lain kesamaan model antara partisipan yang jujur dan menipu adalah strategi yang dilakukan oleh partisipan dalam memanipulasi responnya (Zickar, Gibby, & Robie, 2004). Dari deskripsi statistik didapatkan informasi pendukung bahwa upaya partisipan untuk memberikan impresi positif tidak dilakukan dengan sekedar memanipulasi responnya untuk mendapatkan skor maksimal skala. Pada kondisi motivatif seperti ini, subjek menghindari untuk mendapatkan skor maksimal skala. Hal ini dikarenakan skor maksimal akan

memberikan informasi bahwa mereka memberikan respons yang menipu. Analisis lebih lanjut untuk menjawab mengapa ketiga faktor di dalam skala kepribadian lima faktor rentan terhadap respons tipuan sedangkan kedua faktor lainnya tidak dapat diwujudkan dalam penelitian lanjutan yang menggunakan analisis secara mendetail. Hasil ini didukung oleh hasil penelitian eksperimental yang menggunakan desain mirip dengan desain penelitian ini yang menemukan bahwa respon tidak benar-benar memberikan respon yang menipu sesuai dengan instruksi yang diberikan (Eid & Zickar, 2007).

Penelitian ini menemukan bahwa nilai varians skor antara partisipan pada kondisi netral dan motivatif. Penulis berargumen bahwa jika variasi di dalam respon pada kondisi netral diakibatkan oleh variasi karakteristik kepribadian antar partisipan, variasi di dalam respon motivatif diakibatkan oleh perbedaan strategi dalam memberikan impresi positif. Argumen ini didukung oleh pernyataan Zickar, *et al.* (2004) yang mengatakan bahwa perbedaan skor di antara para pelamar pekerjaan dapat disebabkan oleh perbedaan motivasi, perbedaan tingkat kejujuran, dan kemampuan untuk memalsukan respon. Partisipan yang memiliki motivasi tertentu memiliki persepsi dan kemampuan berbeda terhadap butir-butir di dalam skala.

Temuan penelitian ini yang kedua adalah mengenai probabilitas partisipan untuk mendapatkan skor butir tinggi yang berbeda pada tiap faktor. Adanya probabilitas yang berbeda antar partisipan menunjukkan bahwa pengukuran kepribadian dengan menggunakan skala psikologi yang menggunakan teknik pelaporan mandiri (*self report*) tidak tahan terhadap respons tipuan. Subjek yang memiliki motivasi tertentu dalam merespon alat ukur

memiliki probabilitas tinggi untuk mendapatkan yang skor tinggi. Motivasi tersebut diwujudkan dengan memanipulasi responsnya untuk memberikan impresi positif. Secara umum, penelitian ini menemukan hasil yang tidak konsisten mengenai adanya butir yang terjangkit DIF berdasarkan kondisi partisipan. Sebagian butir terjangkit DIF sedangkan sisanya tidak terjangkit. Hasil penelitian ini paralel dengan temuan penelitian Zickar dan Robie (1999) yang menemukan adanya DIF akan tetapi pada sebagian kecil butir pada alat ukur yang dipakai dalam penelitian mereka.

Penelitian ini memiliki beberapa keterbatasan yang perlu dicatat. Pertama, ukuran sampel yang dipakai dalam pemodelan IRT relatif kecil untuk mengembangkan model yang stabil. Kedua, desain penelitian ini yang menggunakan antar amatan dapat mengakibatkan tidak ditemukannya perbedaan parameter butir pada partisipan yang jujur dan menipu. Rekomendasi yang dapat diberikan adalah agar penelitian ke depan menggunakan desain penelitian campuran yang melibatkan manipulasi dalam dan antar subjek. Hasil ini merekomendasikan kepada penelitian ke depan untuk mengarahkan upaya pendeteksian respon menipu pada skala kepribadian yang bersifat pelaporan mandiri dilakukan pada parameter IRT yang lain, misalnya indeks ketepatan personal (*person fit*). Partisipan yang melakukan tipuan dapat dideteksi dari kecocokan responsnya terhadap skala dengan model IRT (Meade, Ellington, & Craig, 2004). Penggunaan indeks kecocokan personal ini merupakan indeks ketepatan yang juga dapat dipakai sebagai model untuk mendeteksi respon menipu. Indeks ini akan menunjukkan ketidaktepatan pola respon individu terhadap konstruk yang diukur. Respon individu terhadap

butir - butir alat ukur yang tidak konsisten dengan tingkat abilitasnya dapat diprediksi bahwa individu tersebut melakukan respon menipu.

Penelitian ini diselenggarakan atas dukungan dana hibah kompetisi oleh Fakultas Psikologi UGM tahun 2009.

Bibliografi

- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A Meta-Analytic Investigation of Job Applicant Faking on Personality Measures. *International Journal of Selection and Assessment, 14*(4), 317-335.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Counseling Psychology, 24*, 349-354.
- Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996). *The validity of non-cognitive measures decays when applicants fake*. Paper presented at the Annual meeting of the Academy of Management, Cincinnati, Ohio.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Eid, M., & Zickar, M. J. (2007). Detecting response styles and faking in personality and organizational assessments by mixed rasch models (pp. 255-270). New York: Springer.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists: Multivariate applications book series* Mahwah (NJ): Lawrence Erlbaum Associates, Inc.
- Hough, L. M., Eaton, N. K., Dunnette, M. D.,

- Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities [Monograph]. *Journal of Applied Psychology, 75*, 581-595.
- John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. Dalam L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research (2nd ed.)*. New York: Guilford Press.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory - Versions 4a and 5a*. California: University of California, Berkeley, Institute of Personality and Social Research.
- Levine, M. V., & Drasgow, F. (1988). Optimal appropriateness measurement. *Psychometrika, 53*, 161-176.
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics, 4*, 269-290.
- Linacre, J. M. (2000). WINSTEPS: Rasch Model Computer Program (Version 3.68.2). Chicago: Winstep.com.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.
- Meade, A. W., Ellington, J. K., & Craig, S. B. (2004). *Exploratory measurement invariance: a new method based on item response theory*. Paper presented at the Symposium presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology,, Chicago.
- Montag, I., & Comrey, A. L. (1990). Stability of major personality factors under changing motivational conditions. *Journal of Social Behavior and Personality, 5*, 265 – 274.
- Natarajan, V. (2009). *Basic principles of IRT and application to practical testing & assessment*. Bangalore MeritTracers.
- Paulhus, D. L. (1984). Two component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.
- Reise, S. P., & Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Applied Psychological Measurement, 14*, 45-58.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory Factor-Analysis and Item Response Theory - Two Approaches for Exploring Measurement Invariance. *Psychological Bulletin, 114*(3), 552-566.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monographs, 34*((Suppl. 17)).
- Schmit, M. J., & Ryan, A. M. (1993). The big five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78*(966-974).
- Stark, S., Chernyshenko, O. S., Chan, K. Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology, 86*, 943-953.
- Stark, S., Chernyshenko, O. S., Chan, K.-Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology, 86*, 943-953.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological*

Measurement, 59, 197-210.

- von Davier, M. (2010). Why sum scores may not tell us all about test takers. *Newborn and Infant Nursing Reviews*, 10(1), 27-36. doi: 10.1053/j.nainr.2009.12.011
- Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2008). Detection of aberrant responding on a personality scale in a military sample: An application of evaluating person fit with two-level logistic regression. *Psychological Assessment*, 20(2), 159-168.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, 20(1), 71-87.
- Zickar, M. J., & Gibby, R. E. (2006). A history of faking and socially desirable responding on personality tests. Dalam R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior*. Greenwich, CT: Information Age Publishing.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. [doi:10.1037/0021-9010.84.4.551]. *Journal of Applied Psychology*, 84(4), 551-563. doi: 10.1037/0021-9010.84.4.551
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods*, 7(2), 168-190.
- Zickar, M. J., Rosse, J., & Levin, R. (1996). *Modeling the effects of faking on personality instruments*. Paper yang disajikan pada The Annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.